ATTACHMENT B

Responses to EPA: Hestmark letter 8HWM-FF received 10/25/93 and to CDH letter "DOE Proposed Methodology for Statistical Comparison of Remedial Investigation Data at the Rocky Flats Plant" from G. Baughman to R. Schassburger, dated 10/13/93

Response to EPA· Hestmark letter 8HWM-FF received 10/25/93

1 To determine the appropriate background and operable unit populations for comparison, we understand that some matching of the two populations is done by geologists and chemists Data for an analyte in a non-background area are grouped according to a combination of background classes which represent independent background populations A table that cross references the operable unit populations and the background populations will be provided.

> **Concur.** The strawman has been changed to require tables that cross-reference OU media to background media.

2. A more explicit statement of the null hypothesis that is being tested will be included. In addition, a fixed p value of 0.05 will be used for each of the inferential statistical tests as written in the strawman proposal. There was some inconsistency in what was written in the proposal and what was stated in the meeting regarding the p value. A fixed value of 0.05 is what we will accept.

> **Concur.** The strawman states that p values must be less than or equal to 0.05 to demonstrate a significant difference from background. Footnote 3 on page 5 of the strawman, which was not clear on this point, has been deleted.

3 All references to comparison of background and operable unit populations for organics will be removed. Background comparisons apply to inorganics and radionuclides only.

> **Do not concur.** Although background comparisons for organics are not commonly used, there are instances when it may be applicable, in which wide-ranging organic contamination is due to non-site-specific anthropogenic sources We want to retain the option of performing background comparisons for these organics, when geochemists or geologists determine that it is applicable to do so In these instances, we will retain the burden of proof, and the applicability of the comparison will be subject to EPA and CDH approval
>
> The strawman has been rewritten to state that background comparisons for organics will be done on a limited, case-by-case basis, subject to EPA and CDH approval

4 The use of professional judgement in interpreting the results of the graphical displays and statistical analyses will be limited to consideration of spatial distribution, temporal distribution, and pattern recognition concepts The strawman proposal included five additional criteria These will be deleted in the final implementation document.

> **Concur.** The five criteria (intermedia interactions and geochemical processes, not an expected contaminant, blank data, regional background range, and influence of field activities) have been deleted.

parameters. The quantile test could be correctly applied only if the largest n values were all detects. Our statisticians have stated that, typically, this restriction equates to the largest 20% or less of the combined sample sizes being detects, and recommend using a flat 20% to simplify application.

c What is the basis for the criteria of N>20 value for background and operable unit data?

Clarification. Our statisticians derived this value from application of the Central Limit Theorem for a two sample problem. If both samples have N=20, then there will be 38 total degrees of freedom, which will permit assumptions about the distribution.

7. EG&G's claim that these impacts [of implementing Dr Gilbert's recommendations] could range from $30,000 up to $120,000 per operable unit is not supported by the information provided. In fact, it appears that there is some evidence that implementation will not negatively impact costs or schedules

Do not concur. Because the Gilbert method requires additional work, there will be cost and/or schedule impacts.

In addition to the impacts mentioned above, cost impacts may result if the Gehan method is used. For OU11, approximately 200 hours were required to perform the Gehan test, when less than 40 hours would have been sufficient to perform standard ANOVA testing. However, the majority of these costs appear to be one-time costs such as coding development. Subsequent testing on the same OU indicate that the cost impacts may be as little as 30 hours for a small data set.

5. The non-background population is defined as the entire operable unit remedial investigation set. The data aggregation for the purpose of background comparison will be done within the area defined by the operable unit boundaries

Concur. Analysis will be done on an OU-wide basis.

6. The attached flowchart, "Background Comparison Methodology", distributed at the meeting will be clarified. It is EPA's understanding that all the data sets will undergo the hot measurement test and the battery of inferential statistical tests (Gehan, Quantile, Slippage, and T-Test) provided the data satisfies the conditions stated in the strawman and on the flowchart. If any one of these tests, including the hot measurement test, shows significance, the analyte will be further considered, using professional judgement, as a contaminant of concern. The flowchart would benefit from the addition of decision blocks after each test indicating the next step if significance is demonstrated or not.

Clarification. The chart "Background Comparison Methodology" attached to EPA's memo is not the same as that distributed at the September 29, 1993 meeting and contained within the strawman proposal. The difference is that nonparametric ANOVA tests are given as options to the Gehan test in the chart within the strawman proposal. Because the Gehan method is not standard and will therefore incur practical liabilities (e.g., the method has not been adequately tested and verified, preliminary usage shows it to require excessive man-hours, and subcontractors will need to be instructed in its use), we want to retain the option of performing standard nonparametric ANOVA testing, using the Wilcoxon or Kruskal-Wallis tests, instead of the Gehan test.

Additional clarification The suggested decision blocks are not necessary All tests will be performed, if applicable, regardless of whether other tests demonstrate significance.

Concur with the need to redo the flowchart This has been done

6 (continued) We also have some specific questions that need to be addressed in the final document·

a What happens to data which is carried through the slippage test but does not qualify for the t-test?

Clarification. The data that do not qualify for the t-test will be routed to the "At Least One Test Significant?" block The flowchart has been revised to show this

b. What is the basis for the 20% detect value as the criteria for the Quantile test? How does this criteria relate to the criteria for applying this test as stated in Dr. Gilbert's report on page 20?

Clarification Dr Gilbert's method proposed looking up tabulated values for n and r

Response to CDH letter "DOE Proposed Methodology for Statistical Comparison of Remedial Investigation Data at the Rocky Flats Plant" from G. Baughman to R. Schassburger, dated 10/13/93

1. To minimize any potential future misunderstandings of this agreement, the Division feels that it is critical for the Agencies to develop a formal guidance/policy document institutionalizing the agreement. The Strawman document was written for the purpose of facilitating agreement among the Agencies. However, the end users of this document will be the operable unit managers and sub-contractors preparing and reviewing RFI/RI reports. The majority of these people were not involved in the development of this methodology. It is critical to the future of this agreement that final documentation of this agreement be developed to clearly and concisely guide future end users in the implementation of this methodology. This formal guidance should be completed in parallel with the implementation of the agreement.

> **Concur.** When the strawman has been completed and accepted by all concerned parties, it will then be rewritten as a procedure for statistical comparison of OU data to background.

2 The Division recommends that the title of this document be revised to more accurately reflect its content and intent, that being methodology and guidelines for the comparison of site data to background data. The Division proposes the title, "Guide for Conducting Statistical Comparisons of RFI/RI Data and Background Data at the Rocky Flats Plant," for consideration.

> **Concur.** The CDH's proposed title is an improvement to the current title, and has been adopted

3 One of the central themes of Dr Gilbert's recommendations was the need for statisticians to be involved throughout the entire process However, statistician involvement is not discussed in the methodology. The division requests that the role of the statistician in implementation of this methodology be clarified in this document.

> **Concur** Statisticians will be employed to verify that the methods used are correct The strawman has been rewritten to incorporate this

4 The Division does not believe that references to specific DOE sub-contractors are appropriate in this document. The Division recommends DOE review all references to sub-contractors and, where appropriate, modify the reference to more accurately reflect DOE's role and responsibilities.

> **Concur.** References to DOE subcontractors have been eliminated

5 This section (Determine Background and OU Target Populations) outlines the steps for matching site and background populations However, it is unclear exactly how the matching will be implemented. The Division recommends that the rationale for combining media/geology groupings for testing be detailed in this section For example, any criteria for minimum group size necessary for statistical testing should be specified. The Division further recommends adding a table or diagram depicting the general rationale for grouping data by media and geology.

> **Concur.** The strawman states that the OU will match one or more of several specified background media. In addition, the strawman has been changed to require that a cross-reference be performed between the site and one or more background media.

6. As discussed during the September 29th meeting, and emphasized by Dr. Gilbert, it is critical to statistical hypothesis testing that the hypothesis to be tested is explicitly defined and clearly stated. The Division recommends a statement of the test and null hypotheses, in both "english" (narrative qualitative description) and statistical terms, be added to this section of the methodology so there is no misunderstanding of what is being tested. This statement should also address confidence and power requirements for the tests.

> **Concur.** The strawman has been modified to require statistical and prose statements of the null and alternative hypotheses.

7. The Division does not agree with the blanket statement at the beginning of this discussion, "Under current IAG schedule conditions, analytical data will not be 'validated' when the background comparisons will be made in each draft report." This claim is not substantiated by the schedules submitted by DOE in the approved OU work plans and is in direct contradiction to Dr. Gilbert's Task 5 recommendations. Dr. Gilbert states that, "These data quality evaluations are conducted prior to descriptive graphical analyses and formal statistical tests." In finalizing this methodology, the Division recommends that DOE follow Dr. Gilbert's recommendations for data validation before formal graphical presentation and statistical testing The need for variance from this approach will be considered by the Division on an OU specific basis

> **Do not concur.** Under the present system of data validation, the non-validated data are used only for the draft RFI/RI. The final RFI/RI is based solely upon validated data. The lag time between receiving data from the laboratory, and validated data from the independent subcontractor can exceed one month. Waiting for 100% validation may impact schedules, but will probably not change the results in the final RFI/RI. The potential impacts of using non-validated data at each OU will be discussed on a case-by-case basis

8 The Division recommends DOE add a discussion of detection limits to this section of the methodology. In the past there has been confusion as to what detection limits are being reported and used (instrument detection limits vs contract limits vs reporting limits) Part of this confusion may be because detection limits have not been formal discussed This section

should state what detection limits are to be used in statistical testing and how they are determined from the RFEDS data set.

> **Concur.** The strawman addresses detection limits, and it specifies how determinations are made on how to handle non-detects.

9. The Division recommends that this section (Preliminary Exploratory Data Appraisal) be moved to the Data Presentation section.

> **Clarification.** The Data Presentation section consists entirely of deliverables to the EPA and CDH. The preliminary exploratory data appraisal is intended for the use of the analyst only, and does not necessarily constitute a deliverable. For this reason, we have chosen to segregate the two sections.

10. The Division interprets this section as describing the informal data analysis conducted during RFI/RI preparation and not normally included in the formal RFI/RI report. The Division recommends adding language to indicate that this informal data analysis will be made available and reviewed with the regulators in evaluating the appropriateness of the scope of the formal RFI/RI proposal.

> **Do not concur.** We have provided this section for information only. Its products are not intended to be deliverables. If they were to be deliverables, this would impact the schedule of analysis. We have added language to this section to clarify this.

11 The Division does not agree with DOE's recommendations that box plots are applicable only when there are no non-detects. The problem of estimating percentiles for data sets with multiple non-detects was not resolved by Dr Gilbert. The Division recommends that when a reasonably small percentage of non-detects are present, percentiles be estimated using Maximum Likelihood Estimation (MLE) techniques in constructing box plots

> **Concur.** We will provide box plots unless the percentage of non-detects exceeds 50%. The 50% figure is chosen for consistency with the 1993 Background Geochemical Characterization Report (September 30, 1993).

12 The Division does not agree with DOE's suggestion that histograms are not useful for small or highly censored data sets, such as inorganics As stated by Dr. Gilbert, such histograms are not likely to be useful in visually assessing whether the data sets are better modeled by a normal or lognormal distribution However, they may still be useful to visually compare the spread, central tendency, and skewness of the two data sets to look for differences that may be important

> **Concur.** We will provide histograms unless the percentage of non-detects exceeds 50% Bars in the histogram will be shaded to indicate the percentage of detects and non-detects within each bar interval.

13 The Division recommends that a discussion be added to this section of the methodology

to address what to do when a UTL 99/99 can not be reasonably estimated or is unknown (ie small or highly censored background data set).

Concur. We have modified the strawman to state that professional judgement, and use of geochemical background data from the literature, will be used The result will be a geochemical interpretation of data, subject to agency review and approval.

14. The reference in Footnote 2 to OU 1 is not appropriate and should be removed. The inferential tests conducted at OU 1 were the result of a compromise agreement, are not precedent setting for other OUs and are not the tests being proposed in this document. However, as stated in this note, limited professional judgement as presented later in this document may be applicable.

Concur. This footnote has been deleted.

15. This discussion (Footnote 3) should be moved to the DQOs or statistical test definition section of the document.

Clarification. This footnote has been deleted. We intend to use a p value of 0 05, and the footnote made that intent unclear.

16. The Division does not agree with the limitations DOE has placed upon the Slippage Test. The slippage test can be applied to data sets when the largest background point is a non-detect If the largest background data point is a non-detect then logic must be applied to determine if the slippage test is applicable, but the test should not be categorically eliminated.

Concur. We have rewritten the strawman to state that, if the largest background data point is a non-detect, we will apply judgement to investigate whether or not the slippage test is applicable.

17. The Division recommends limiting the use of professional judgement to the first three criteria, spatial distribution, temporal distribution, and pattern recognition. In addition, it is recommended that the introduction to this section include acknowledgement that in applying professional judgement, the "burden of proof" lies solely on DOE Professional judgement will only be considered by the Division on a limited basis where well documented and defensible evidence is presented.

Concur. We have eliminated the last five criteria from the strawman, and acknowledged that we will bear the burden of proof

18. To make the process more efficient the task of eliminating non-detected analytes should be completed prior to data presentation. The flow chart should be modified to reflect this change.

Concur We have changed the flowchart CDH's comment improved the process

19 This flow chart is confusing and difficult to follow due to the many multiple and undefined branches To minimize the potential for misunderstanding this chart must either be clarified or deleted.

Concur. The flowchart is too important to delete. It has been clarified Lines denoting the flow of information have been deleted, keeping only the lines denoting flow of control, in accordance with common flowcharting techniques. Decision blocks have been transformed into diamond shapes Alternative "No" paths have been added for the blocks labeled "No Non-Detect Present ..OU Data Normally Distributed?", and "At Least One Test Significant?" Finally, the block representing the conditions which must be met prior to performing the t-test has been changed to reflect the conditions given in the text.